

Grade 9 Social Studies Achievement Testing Program

Provincial Report

October 1983

Distribution:

Officials of Alberta Education

School Principals

Superintendents of Schools

The Alberta Teachers' Association

Alberta School Trustees' Association

General Public upon Request



EXECUTIVE SUMMARY

Organization of the Test

The Grade 9 Social Studies Achievement Test consists of two sections: a multiple-choice section weighted 70%, and a written-response section weighted 30%. The multiple-choice section consisted of 60 questions that were machine scored, and the written-response section contained five questions that were scored by 84 teachers of Grade 9 social studies under the supervision of Alberta Education staff.

It was determined that an acceptable provincial standard of achievement on the test would be for 85% of students to receive marks of 40% or higher.

Administration

The test was administered to more than 23 500 students enrolled in Grade 9 social studies on June 14, 1983. Students in two jurisdictions were selected for testing by random sample of classes or schools; in all other jurisdictions, all students currently enrolled were tested.

Findings

The provincial average on the test was 59.2%. The average for the 60 multiple-choice questions was 36.6 (61.0%). The average for the written-response questions was 16.5 (55.0%). Since 87.8% of students received marks of 40% or higher, achievement on the test was judged to be acceptable.

TABLE OF CONTENTS

	PAGE
PREFACE	v
ACKNOWLEDGMENTS	v
CHAPTER 1: THE ACHIEVEMENT TESTING PROGRAM	1
Exemptions from the Achievement Testing Program	1
CHAPTER 2: TEST DEVELOPMENT, DESCRIPTION, AND DESIGN	2
Test Development	2
Test Description	2
Multiple Choice	2
Reporting Categories	4
Written Response	12
Determination of the Student Population	15
Administration	15
Data Collection	15
Test Design Statistics	16
CHAPTER 3: SCORING OF WRITTEN-RESPONSE SECTION	17
CHAPTER 4: RESULTS	22
Guide to Interpretation	22
Absentee Rates	24
Standard Setting	25
Results	26
CHAPTER 5: CONCLUDING OBSERVATIONS	31

LIST OF FIGURES

FIGURE		PAGE
1	Grade 9 Social Studies Achievement Test Blueprint of Multiple-Choice Section	3
2	Reporting Categories	4
3	Sample Multiple-Choice Questions	6
4	Excerpts from Written-Response Scoring Guide	18
5	Scoring Guide for Question 5	21
6	Target Regions for Grade 9 Social Studies Achievement Test Multiple-Choice Section	23
7	Target Ranges for Grade 9 Social Studies Achievement Test Written-Response Section	24

LIST OF TABLES

TABLE		PAGE
1	Total Test Results	27
2	Multiple-Choice Section Results	27
3	Written-Response Section Results	27
4	Written-Response -- Question 5	28
5	Item Response Frequencies	29

PREFACE

This report presents the provincial results of the Grade 9 Social Studies Achievement Test administered June 14, 1983, as part of the student achievement testing program of Alberta Education. More than 23 500 students wrote the test. This report is intended to provide more information about the test, the test development process, and the results of both the written-response and the multiple-choice sections. It is also written to assist local school boards and administrators in interpreting the results for their jurisdictions.

ACKNOWLEDGMENTS

The successful administration of the Grade 9 Social Studies Achievement Test was the result of the concerted effort of all involved. Success would have been impossible without substantial contributions from many people, particularly the students, teachers, and administrators who extended their full co-operation.

The technical expertise of Dr. T. O. Maguire, Professor, Division of Educational Research Services, University of Alberta, has been particularly valuable in the implementation of the Achievement Testing Program. His contributions to the design, analysis, and reporting are gratefully acknowledged.

Lloyd E. Symyrozum
Director
Student Evaluation Branch

Chapter 1

THE ACHIEVEMENT TESTING PROGRAM

The Achievement Testing Program is designed to provide extensive and reliable indicators of levels of student achievement in basic education throughout the province. The tests are specific to the *Program of Studies* prescribed by the Minister of Education. The curriculum specifications, which identify the instructional emphases to be placed on the content and objectives of the curriculum, serve as the basis for each of the achievement tests. These tests provide data that facilitate the establishment of provincial standards for the long-term appraisal of student achievement.

The program tests levels of student achievement in science, mathematics, social studies, and the language arts in Grades 3, 6, and 9. Not all tests are given in each grade each year; rather, the subject areas are tested on a cyclical basis. In 1983, achievement tests were administered in Grade 3 science, Grade 6 mathematics, Grade 9 social studies, and Physics 30.* The provincially set achievement tests are designed to yield results meaningful at both the jurisdictional and provincial levels.

Following each test administration in June, the results are reported in the fall of the next school year. A *Summary Report* that presents the highlights of the test administration is prepared for public distribution. This report, the *Provincial Report*, is more extensive. It is designed to assist jurisdictions in interpreting the results. Each jurisdiction also receives summary tables for that jurisdiction, including results listed by school and students. Alberta Education does not issue individual statements of results to students.

Exemptions from the Achievement Testing Program

Under normal circumstances, the following classes are exempt from achievement testing:

- Special Education classes registered for grants with the Special Educational Services Branch
- classes in which the language of instruction is other than English, under the terms of Section 159 of the *School Act*
- classes in which the subject being tested has been cycled and taught in an alternate year
- classes in which the subject has been taught in a semester other than that in which the test is being administered
- classes for students for whom English is a second language

* After 1983, achievement testing will not be conducted at the Grade 12 level because of the introduction of the Grade 12 Diploma Examinations.

Chapter 2

TEST DEVELOPMENT, DESCRIPTION, AND DESIGN

Test Development

There were three stages in the development of the Grade 9 Social Studies Achievement Test: preparation of curriculum specifications, development of questions, and selection of questions for the final test.

The Curriculum Branch of Alberta Education prepared curriculum specifications that identified the major content areas, the specific objectives within each area, and the emphasis each is to receive in the classroom. The curriculum specifications were distributed to all school jurisdictions in the province, and teachers were invited to react to them.

Under the supervision of the Student Evaluation Branch of Alberta Education, test questions were written, field tested, and revised by classroom teachers. Questions were designed to reflect the value, knowledge, and inquiry skill objectives outlined in the curriculum specifications. Although attitudes and participation skills are included in the specifications, these curricular areas were not tested.

The final test was constructed from those questions that were approved by the teacher committees. The draft test was submitted to Potenz Editorial, specialists in determining readability, to ensure that the test was at a reading level of Grade 9 or lower. The final draft was examined by a Technical Review Committee consisting of two teachers from the subject area being tested, representatives from several stakeholder groups (The Alberta Teachers' Association, Conference of Alberta School Superintendents, and universities), and representatives of the Curriculum Branch, the Student Evaluation Branch, and the Regional Offices of Alberta Education. The Technical Review Committee checked the draft for content validity, accuracy, and technical merit, and changes recommended by the Committee were made.

Test Description

The Grade 9 Social Studies Achievement Test consists of two sections: a multiple-choice section weighted 70% and a written-response section weighted 30%.

Multiple Choice

The multiple-choice section of the test consists of 60 items. Emphasis in terms of attention to value, knowledge, and skill objectives is 12%, 44%, and 44% respectively.

Items based on the value objectives in the multiple-choice section relate to understanding values. Skills in value analysis, decision-making, and moral reasoning are evaluated in the written-response section. The objectives related to the development of attitudes are not included in this test; the 10% weighting allocated to attitudes in the curriculum specifications was reassigned to the remaining objectives on a prorated basis.

Items based on skill objectives relate to the inquiry skill categories.

Items based on knowledge objectives are drawn equally from the three topics prescribed for Grade 9.

Figure 1 provides a detailed test blueprint showing the distribution of items across topics and objectives.

Figure 1
Grade 9 Social Studies Achievement Test
Blueprint of Multiple-Choice Section
(Distribution of Items among Topics and Objectives)

Objectives	Topic A	Topic B	Topic C	Total Number of Items
Value				
- Understand Values	3	3	2	8
- Develop Competencies				
Subtotal	3	3	2	8
Knowledge (Concepts)				
- Basic Economics		1		1
- Centrally Planned Economy				
- Historical Evolution		3		3
- Principles of Centralization		2		2
- Conservation			1	1
- Control		2		2
- Demography			1	1
- Industrialization				
- In Britain	4			4
- In Canada (Geog. Factors)			1	1
- Industry				
- Primary, Secondary, Service			3	3
- Labor/Management Relations	1			1
- Market Economy	1			1
- Materialism	1			1
- Quality of Life	1			1
- Scarcity	1			1
- Technological Change			2	2
- Welfare of the State		1		1
Subtotal	9	9	8	26
Skill				
- Inquiry				
a. Identify and Focus on Issue	3	1	2	6
b. Formulate Research Questions	1	2	1	4
c. Gather and Organize Data	3	3	1	7
d. Analyze and Evaluate Data		2	1	3
e. Synthesize Data	1		2	3
f. Resolve the Issue			2	2
g. Apply the Decision			1	1
h. Evaluate the Decision				
Subtotal	8	8	10	26
Total Number of Items	20	20	20	60

Reporting Categories

To provide information that is both meaningful and reliable, it is necessary to group items into reporting categories (subtests). Each reporting category requires a minimum of six items to achieve statistical reliability. Figure 2 indicates the reporting categories used in the Grade 9 Social Studies Achievement Test and the number of items in each category. (Note that items may appear in more than one reporting category.)

Figure 2
Reporting Categories

1. Topic A: Selected Market Economies	20 items
2. Topic B: Selected Centrally Planned Economies	20 items
3. Topic C: Industrialization in Canada	20 items
4. Recalls basic economic, historical, and geographical facts related to industrialization	8 items
5. Recalls facts about industry and technology in Canada	7 items
6. Understands key ideas in economics, history, and geography, specifically those related to economic systems and technological change	11 items
7. Understands underlying values and key issues related to industrial growth	8 items
8. Can identify major issues and recognize differing viewpoints on industrial growth	9 items
9. Can select appropriate research questions and apply effective ways of gathering and organizing information	11 items
10. Can analyze, evaluate, and synthesize information related to a topic under study	9 items

Questions were arranged on the test by curriculum topic rather than by reporting category. A serious attempt was made to simulate the inquiry process approach even within the multiple-choice section of the test. Students were presented with "families" of items related to a single issue or data source. Within each family of items -- and as much as possible between families -- students followed a logical progression from one question to the next such that they would be required to identify an issue, identify the underlying value positions or conflict, recall appropriate knowledge, and/or apply their inquiry skills to data. Individual questions were integrated into a logical flow through each topic similar to the sequence followed in the inquiry process. Psychometric considerations -- especially the need to ensure that each item could be answered correctly, whether or not the student had answered previous items in the sequence correctly -- severely limited the degree of integration that could be achieved. Nevertheless, it was felt that this approach was superior to presenting students with a series of completely unrelated items, or arranging questions by subtests (that is, all value questions, followed by all knowledge questions, and then all skill questions).

Figure 3 illustrates a typical family of multiple-choice questions from the 1983 Grade 9 Social Studies Achievement Test. All of these items are from Topic C: Industrialization in Canada (Reporting Category 3). Items 46-50 are also reported in Category 5: Recall facts about industry and technology in Canada. Items 42 and 43 are from Category 8: Can identify major issues and recognize differing viewpoints on industrial growth, with item 43 also reported in Category 7: Understands underlying values and key issues related to industrial growth. Items 44 and 45 are examples of Category 9: Can select appropriate research questions and apply effective ways of gathering and organizing information. Items 51, 52, and 53 are from Category 10: Can analyze, evaluate, and synthesize information related to a topic under study.

Figure 3

Sample Multiple-Choice Questions

SOURCE IV



42. One major issue identified in these sources is:
- A. Should multinational oil companies be the only ones allowed to develop Arctic oil resources?
 - B. Should Arctic oil development involve the cheapest methods possible?
 - C. Should our trade relations with Middle East nations affect Canada's oil exploration and development?
 - D. Should Arctic oil exploration and development be encouraged?
43. The values in conflict between the authors of Sources I and III are
- A. self-sufficiency and economic security
 - B. environmental concern and national unity
 - C. human dignity and self-sufficiency
 - D. environmental concern and economic security

Figure 3 (Continued)

44. To find out how an energy shortage would affect the quality of life of Canadians, the most useful research question would be:
- A. To what extent do jobs and economic growth depend upon an adequate supply of energy?
 - B. Can a practical electrical car be designed to replace gasoline-powered automobiles?
 - C. To what extent are oil supplies threatened by continued political unrest in the Middle East?
 - D. Can the quality of Alberta petroleum products meet world market standards?
45. According to the map in Source IV, the regions with the greatest potential for oil and gas development in Canada are the
- A. Great Central Plain and Cordillera
 - B. Cordillera and Canadian Shield
 - C. Canadian Shield and Arctic Lowlands
 - D. Arctic Lowlands and Continental Shelf
46. If closeness to raw materials were the only factor to be considered in locating secondary industries, Canada's plastics industry would currently be located in
- A. Ontario
 - B. Alberta
 - C. Nova Scotia
 - D. Quebec
47. If industrial development is to take place in a remote area, special steps must be taken to
- A. attract workers
 - B. improve the roads
 - C. pipe in a water supply
 - D. establish chain stores
48. The major geographic factor limiting industrial growth in Canada's North is the
- A. harsh climate
 - B. shortage of mineral resources
 - C. unproductive soil
 - D. lack of suitable harbors

Figure 3 (Continued)

49. An example of a service industry that would be involved in the construction of an Arctic pipeline would be the
- A. bank that arranged financing for the project
 - B. factory that produced the pipe
 - C. construction company that built the pumping stations
 - D. company that built the construction equipment
50. Differences of opinion about how to develop the petroleum industry in Canada have caused a serious dispute between
- A. the provincial governments of British Columbia and Ontario
 - B. the federal government and some provincial governments
 - C. Canada and the other members of the Commonwealth
 - D. the provincial governments of Newfoundland and Alberta
51. A broad goal of Canadian society that would be advanced by the development of Arctic gas and oil reserves is the
- A. achievement of energy self-sufficiency
 - B. urbanization of the Canadian North
 - C. preservation of the Inuit culture
 - D. improvement of federal-provincial relations
52. The most likely result of discouraging Arctic exploration for oil and natural gas reserves would be
- A. increased employment
 - B. reduced oil imports
 - C. protection of the environment
 - D. increased corporate profits
53. Of the following, the best way of assessing the impact of Arctic pipeline construction on a community in Alberta would be to
- A. make a chart showing energy costs, rates of unemployment, and environmental problems in the community before and after the pipeline is built
 - B. write a letter to the head of the pipeline company and ask for information on problems involved in the construction of the pipeline and how they were solved
 - C. write a letter to the head of an environmental protection group and another to a native cultural group and ask for their views on the pipeline construction
 - D. make a chart showing how much oil was pumped through the pipeline in the first year of operation and estimate the total number of barrels of oil that will be transported over the lifetime of the pipeline

Figure 3 (Continued)

Use the information below to answer questions 54 to 57.

Assume that the federal government has decided to develop natural resources on several large parcels of Crown land located in Canada's North. Various oil and mining companies will be invited to submit bids for claims. Below are four responses that might greet a government announcement that it is going ahead with northern development.

SPEAKER W: The money gained through taxes and royalties will help to reduce Canada's budget deficit substantially and help build highways or pay for medical care. Canada's resources are there to be used.

SPEAKER X: Canada's North is a treasure in trust. We must preserve this wilderness area for future generations of Canadians. Development is not worth the price of a damaged environment.

SPEAKER Y: The Canadian economy can best be strengthened through industrial growth. Mineral resources are worthless unless developed. Think of all the jobs that will be created.

SPEAKER Z: I've been laid off from my job on the oil rigs for over a year now, and am in danger of losing my home because I can't make the mortgage payments.

54. The main concern expressed by Speaker W is the need for the government to
- A. create employment
 - B. raise additional money
 - C. create industrial growth
 - D. preserve the environment
55. The values in conflict between Speaker X and Speaker Z are
- A. materialism and right to work
 - B. materialism and industrial growth
 - C. conservation and wealth
 - D. conservation and job security
56. If Speaker Y's solution for curing Canada's economic problems were carried out, Speaker X would be most concerned about the possibility that
- A. Canada's economy would become increasingly resource-based
 - B. Canada's economy would become too dependent upon foreign companies
 - C. many hectares of wilderness would be permanently polluted or destroyed
 - D. many important development projects in southern Canada would be dropped in favor of northern development

Figure 3 (Continued)

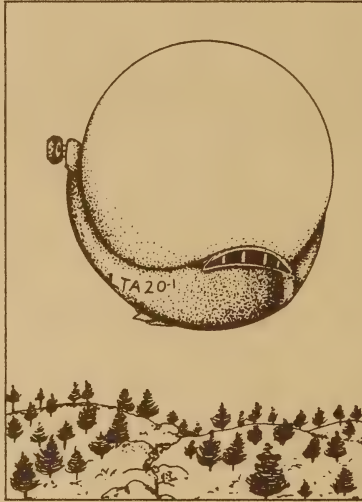
57. The majority of the speakers evaluate resource development in terms of
- A. environmental needs
 - B. energy needs
 - C. economic needs
 - D. local employment needs
-
58. One effect of technological change on the Canadian labor scene is that
- A. the time required to complete most tasks has increased
 - B. working conditions have generally become hazardous
 - C. manual labor in primary industries has increased
 - D. many occupations no longer exist

Figure 3 (Continued)

Use the information below to answer questions 59 and 60.

An Idea Whose Time Has Come Again

Lighter-than-air craft are making a comeback. Modern technology has produced new materials and designs that once again make airships practical and safe. The modern airship can carry between 40 and 70 tonnes of freight at a fraction of the cost of current methods. Everyone agrees that airships are ideal for northern development because they do not require runways (as do airplanes), tracks (as do trains), or roads (as do trucks), all of which are expensive, difficult, and time-consuming to build. All an airship needs to "dock" in a remote area is a portable mooring mast and a large open area. Thus damage to the surrounding wilderness is kept to a minimum, since nothing is disturbed except the actual drilling, logging, or mining site.



Artist's sketch of a new type of airship being developed in Canada

59. According to the article above, the development of the modern airship would be an example of a technological change that will
- A. encourage conservation of our northern oil, forest, and mineral resources
 - B. encourage protection of the environment
 - C. decrease employment in primary industries
 - D. increase our dependence on nonrenewable resources
60. An example of bias in the article is the statement that
- A. "Everyone agrees that airships are ideal for northern development"
 - B. "they do not require runways (as do airplanes)"
 - C. "All an airship needs . . . is a portable mooring mast and a large open area"
 - D. "damage to the surrounding wilderness is kept to a minimum"

Written Response

The written-response section of the test provided students with the opportunity to apply what they had learned to new situations and issues. All the written-response questions were related to a single issue and students were required to work through the steps of the inquiry approach to complete the assignment. The written-response section included value, knowledge, and skill objectives, with particular emphasis on those objectives not easily addressed through multiple-choice questions.

Students were presented with the issue: "Should the introduction of computer technology in Alberta industry be controlled?" A brief statement of the "Background to the Issue" was followed by "Position Summaries," or position statements of four speakers. The students were asked the following questions.

Question 1 asked students to identify conflicting values.

- 5 marks 1. A person's views are usually based on the values held by that person. Choose one of the four speakers from page 26 and name the value that appears to be important to that speaker. Give evidence from the speaker's statement in the position summaries to support your answer.

a. Speaker _____

Value _____

Evidence _____

Choose a second speaker whose views seem to be in conflict with those of the speaker you chose above. Name the value that appears to be important to that speaker. Give evidence from the speaker's statement in the position summaries to support your answer.

b. Speaker _____

Value _____

Evidence _____

Question 2 asked students to recall an appropriate historical event and to evaluate its impact on the quality of life.

- 2 marks 2. a. Give one POSITIVE effect of the introduction of new technology on life in Great Britain during the Industrial Revolution. Give a reason for your answer.

- 2 marks b. Give one NEGATIVE effect of the introduction of new technology on life in Great Britain during the Industrial Revolution. Give a reason for your answer.

Question 3 asked students to analyze and evaluate data: four newspaper extracts, an editorial cartoon, and an elaborate bar graph.

- 2 marks 3. a. From the data on page 29, predict one possible POSITIVE effect of the introduction of computer technology in Alberta industry today. Give a reason for your answer.

- 2 marks b. From the data on page 29, predict one possible NEGATIVE effect of the introduction of computer technology in Alberta industry today. Give a reason for your answer.

Question 4 asked students to synthesize data and produce a generalization based on both historical events and current trends.

- 2 marks 4. From your answers to questions 2 and 3, what generalization (general conclusion) can be made about the effects of the introduction of any new technology on society?

Question 5 asked students to produce an extended response, several paragraphs in length. Students were asked to take a position on the issue and support it. This question required students to analyze and synthesize data, resolve the issue, apply the decision, and communicate effectively.

- 15 marks 5. Imagine that you live in a town whose mayor is thinking of buying a computer system for the civic centre. The new computer would help modernize the town's services and reduce costs. Although taxes would have to be raised slightly for two years to pay for the computer, the long-term savings would allow town council to cut taxes in 1985 or to introduce new services. The computer, however, may take the jobs of many civic workers such as clerks and typists. The mayor has asked the citizens for their opinions on the issue.

Write a letter to town council in which you state whether or not the town should buy the computer. Give reasons for your opinion. Make a convincing presentation. Use evidence from the data provided and from your knowledge of social studies to support your arguments. Sign your letter "Pat Smith."

Your letter will be marked according to the following:

- how persuasive or convincing your letter is
- how accurate and logical your arguments are
- how well you express yourself (grammar, word usage, punctuation)

Determination of the Student Population

Each school jurisdiction could choose to test either all Grade 9 social studies students or only randomly selected classes. To ensure a very small sampling error for the jurisdiction results, however, it was necessary for small jurisdictions to test almost every class. For example, a jurisdiction with 30 classes in a given course would need to test 28 of them to reach the desired confidence level of 2% with 99% confidence. School boards were required to notify the Student Evaluation Branch of their wish to have all of their students write. The two largest jurisdictions, Calgary School District #19 and Edmonton School District #7, opted for sampling. All other jurisdictions required all of their students to write the tests.

Complete lists of eligible Grade 9 classes being taught in the Edmonton and Calgary school districts at the time of the test were obtained. A random sample of classes was drawn from the Edmonton classes. At the request of the Calgary Board of Education, the Calgary students to be tested were selected by a random sampling of schools, stratified by school enrolments. Enrolment in classes or schools selected for testing represented 35% of the eligible students in Edmonton School District #7, and 49% of eligible students in Calgary School District #19.

In some jurisdictions, junior high school instruction is carried out under the semester system. Only students who were enrolled in Grade 9 social studies in June 1983 were eligible to be tested. Those students completing the course in the first semester were not eligible to write.

Administration

Jurisdictions were requested in April to report the number of students enrolled in Grade 9 social studies in each school. In May, letters were sent by the Student Evaluation Branch to superintendents, principals, and teachers in the province requesting their co-operation in the testing. Information addressed to the superintendents and principals included the test schedule, procedures for test administration, and requirements for returning test materials. Information addressed to the teachers related to the administration of the test and the return of test materials. Each jurisdiction was sent the appropriate number of tests and administration instructions, packaged according to school. After the test was administered, teachers were instructed to collect all test booklets and answer sheets and return them to the principal for forwarding to school board offices, which, in turn, were responsible for sending the test booklets and answer sheets to the Student Evaluation Branch.

Data Collection

A total of 23 639 students in 564 schools from 104 jurisdictions returned scorable booklets.

Test Design Statistics

Standard test statistics indicated that the multiple-choice section of the test had levels of reliability suitable for providing data to compare achievement between a jurisdiction and the province. The Kuder-Richardson Formula 20 reliability (internal consistency) coefficient was .888. The intra-test correlations between the multiple-choice and written-response sections was 0.614. This indicates that the two sections were testing related but distinct abilities, and that it is important to include both types of evaluation on the Achievement Test for Social Studies.

Eighty-four classroom teachers were asked to rate the curricular importance of each item on the multiple-choice section to verify that that test was a valid instrument. Modal ratings on the four-point scale were

(1) Questionable	0 items
(2) Acceptable	0 items
(3) Important	50 items
(4) Essential	10 items

SCORING OF WRITTEN-RESPONSE SECTION

A scoring centre was established in Edmonton and 84 markers came from across the province. Markers had to have a valid permanent Alberta teaching certificate, have taught Grade 9 social studies for at least three years, and be currently teaching Grade 9 social studies. The markers were provided an opportunity to review and discuss answer keys prior to marking, and some minor revisions to the keys were made in light of this feedback from practising classroom teachers. Markers were trained and supervised by Alberta Education staff who were available if markers experienced difficulties or had questions.

Before papers were delivered to markers, student identification information was removed and the compositions were placed in bundles of 25. When marking any composition, the scorers had no way of knowing the identity of the student who wrote it, or the student's school.

For marking purposes, each paper was divided into three sections, each section to be scored by a different reader. Thus, while each section was read only once, each paper was scored by three different markers.

This procedure produces results that are almost as reliable at the jurisdictional level as if each paper had been read from beginning to end by three different scorers. If an essay is to be marked reliably at an individual level (for example, to be used for promotion purposes for the individual student), it is highly desirable to have it scored by at least three independent readers to ensure that the student does not get a marker who scores consistently higher (easier) or lower (harder) than average in spite of training and scoring procedures designed to limit these variations. However, at the jurisdictional level where individual scores are not of primary concern, three readers scoring successive questions produces statistically comparable results. In other words, with three readers scoring successive questions, those who are consistently high or low markers cancel each other out every three papers. This procedure represents a very considerable saving in effort and cost.

In addition, reliability reviews were used in training sessions and every half-day in marking Question 5. The 84 markers were divided into 17 groups and a group leader was appointed in each group to chair the reliability reviews. At these reviews all markers were given identical sets of four papers. They independently marked each paper and then compared and discussed results, thereby developing a group rationale for scoring. Results of these group sessions were compiled and distributed to markers at the next reliability review so that markers could compare the scores they had assigned with those most frequently assigned by all the other markers. Thus, consistency of scoring was encouraged among all markers.

Figure 4 presents excerpts from the scoring guides for Questions 1 through 4. Figure 5 presents the scoring guide for Question 5.

Following completion of the scoring, markers were asked to rate the written-response scoring system. The key for Questions 1 through 4 was most commonly rated "generally explicit," and the descriptors for Question 5 were most commonly rated "readily applicable to most of the papers." A majority of the teachers found the training to be useful, with 83% rating the use of sample papers, reliability reviews, and overall training as "very helpful" or "extremely helpful."

Figure 4

Excerpts from Written-Response Scoring Guide

Question 1. One mark for selection of speakers with conflicting views.
One mark for identification of value for each speaker (total of two marks).

One mark for presentation of appropriate evidence for each speaker (total of two marks).

Question 2.A. One mark for stating a positive effect of the introduction of new technology.

A second mark for a valid reason supporting the selected positive effect.

The reason should state, explicitly or implicitly, what value was enhanced, or how the positive effect improved the quality of life. Reference to the value or positive effect should be made within the appropriate historical context.

EXAMPLES:

i. Positive Effect: Increase in Standard of Living

Reason: An ever-increasing supply of goods that were reasonably cheap and of consistent quality meant that more goods were available to more people than ever before.

ii. Positive Effect: Growth and Improvement of Transportation

Reason: Canals were built or enlarged. Railroads and highways were constructed. Ships were improved. Together, these facilities moved more goods and more people farther and faster than had ever before been thought possible.

Question 2.B. One mark for stating a negative effect of the introduction of new technology.

A second mark for a valid reason supporting the selected negative effect.

The reason should state, explicitly or implicitly, what value was undermined, or how the negative effect adversely affected the quality of life. Reference to the value or negative effect should be made within the appropriate historical context.

EXAMPLES:

i. Negative Effect: Creation of Slums

Reason: People flocked to the cities from the country districts to find jobs in industry. As a result, there was usually a shortage of housing for workers. Landlords were able to charge high rents for very poor accommodations. These industrial areas rapidly became dreadful slums.

ii. Negative Effect: Child Labor

Reason: Five- and six-year-old orphans, as well as children of the poor, were forced to work in the factories and coal mines for only a few cents a week. In addition to being very harmful to the health and well-being of the children, child labor also resulted in a surplus of labor, causing reduced wages and lay-offs for adult workers.

Question 3.A. One mark for predicting a possible positive effect of the introduction of computer technology in Alberta industry today.

A second mark for a valid reason supporting the selected positive effect.

The reason should state, explicitly or implicitly, what value would be enhanced or what need fulfilled by the possible positive effect. Evidence supporting the reason should be derived from the Data Page.

Question 3.B. One mark for predicting a possible negative effect of the introduction of computer technology in Alberta industry today.

A second mark for a valid reason supporting the selected negative effect.

The reason should state, explicitly or implicitly, what value would be undermined or what need thwarted by the possible negative effect. Evidence supporting the reason should be derived from the Data Page.

Question 4. Two marks for a clearly stated generalization that meets the following criteria:

1. It shows how two or more concepts are meaningfully related to each other.
2. It is accurate, logically consistent, and directly related to the answers in questions 2 and 3.

One mark awarded for a generalization that is not directly related to the answers in questions 2 and 3.

EXAMPLES:

- The introduction of new technology is often accompanied by both social costs and social benefits.
 - The introduction of new technology affects our quality of life.
 - Increased production of goods may lead to increased concern with materialism.
-

Question 5 requires students to take a stand on a facet of a current social issue as it manifests itself in a specific situation. They are expected to communicate their position effectively. This type of assignment is ideally suited to the primary trait method of scoring. Primary trait scoring was originally developed by the National Assessment of Educational Progress in the United States in the early 1970s. Ina V. S. Mullis describes the underlying rationale for primary trait scoring as follows:

The rationale underlying primary trait scoring is that writing is done in terms of an audience and can be judged in view of its effects upon that audience. Particular writing tasks require particular approaches if they are to be successful. The approach used by the writer to reach and affect his audience will be the most important -- the primary trait -- of a piece of writing. For example, the writer of a set of directions must present things in a logical and unambiguous manner if he expects readers to follow the directions. Therefore, the primary trait of a written set of directions would be an unambiguous, sequential, and logical progression of instructions. Successful papers will have that trait, unsuccessful papers will not, regardless of how clever or well-written they may be in other respects. On the other hand, the purpose of campaign literature is to persuade a reader to vote for a candidate. A successful campaign paper will have certain persuasive traits that an unsuccessful one will not have, and these traits will differ from those necessary for a successful set of directions.

Question 5 asks "Pat Smith" to write a letter to Mayor Applegate, commenting on the advisability of adopting computerization. The best letters, presumably, will be those that persuade Mayor Applegate to adopt the writer's point of view. In other words, the primary trait in this case is persuasiveness. For purposes of scoring, you are asked to put yourself in Mayor Applegate's place. Accordingly, letters that move you, that persuade you most, will receive the highest scores. Letters that are not persuasive cannot receive a high score regardless of how well written they may be in other aspects (e.g., in terms of word choice, sentence structure, or organization).

Figure 5

Scoring Guide for Question 5

Descriptor	Score	Criteria for Scoring
EXCELLENT	5	The writer presents an excellent argument based on solid reasoning and carefully selected facts. His use of language is accurate and effective. His letter is very persuasive and he uses outside information to support his position.
GOOD	4	The writer presents a good argument based on logical reasoning and relevant facts. His use of language is clear and understandable, although there may be a few minor grammatical and spelling errors. His letter is persuasive and he supports his position well.
ACCEPTABLE	3	The writer presents an identifiable position based on related facts. His use of language is adequate, although there may be some grammatical and spelling errors. His letter is somewhat persuasive.
LIMITED	2	The writer presents a confused statement of position. His arguments tend to rely on emotion and not logic. The facts that are cited are only remotely related. His use of language is limited and displays frequent grammatical and spelling errors. His letter is unpersuasive.
POOR	1	The writer does not present a definable position, or he may state a position but makes little or no attempt to defend it. Content is inaccurate or inappropriate. His use of language is seriously flawed with numerous grammatical and spelling errors. His letter either lacks a sense of purpose or is completely unpersuasive.
(0) - "Zero" is a special category. It is not an indicator of quality. It should be assigned to papers that are blank, totally illegible, or address a completely different topic.		

Chapter 4

RESULTS

Guide to Interpretation

The results of the Grade 9 Social Studies Achievement Test are most useful in comparing achievement in a particular jurisdiction with the provincial targets. However, care must be exercised in making these comparisons.

Because many factors influence the performance of students within a jurisdiction, setting provincial targets is a difficult task. Among the many factors that influence the results are the following:

1. The tests have been designed to sample the Grade 9 social studies curriculum. The amount of testing time limits the extent to which it is possible to test the objectives in the curriculum.
2. Whenever achievement is assessed, there are errors of measurement. These arise from many sources and can never be entirely eliminated. Care has been taken to develop highly reliable scoring procedures for the written-response section, but perfect accuracy is not possible.
3. The size of the jurisdiction relates to the stability of results from one year to the next. Average levels of ability are much more likely to fluctuate in smaller jurisdictions than in large ones, with the consequence that the average achievement scores attained by smaller jurisdictions are likely to fluctuate as well.
4. The motivation levels of the students greatly influence the achievement scores.
5. As with any pencil-and-paper test, scores will be affected by reading skills as well as by the skills that the test is intended to measure.

For the purposes of interpreting the results, the target mean was established for the multiple-choice section of the test at 60%, or an average raw score of 36. This target mean was selected after a thorough examination of the results and after careful consideration of the difficulty level of the test. This mean is intended to serve as an indication of a reasonable level of achievement for a typical jurisdiction.

As noted above, jurisdictional results are subject to various influences that tend to reduce their accuracy. To accommodate this, target means are translated into regions or performance bands so that individual jurisdictions can assess their own performances. Figure 6 shows the target regions for the total multiple-choice section and its reporting categories.

Figure 6

Target Regions for Grade 9 Social Studies Achievement Test
Multiple-Choice Section

Reporting Category	Number of Items	Target Region (Average Raw Score)
Total Multiple Choice	60	32-42
Topic A: Selected Market Economies	20	10-14
Topic B: Selected Centrally Planned Economies	20	10-14
Topic C: Industrialization in Canada	20	10-14
Recalls basic economic, historical, and geographical facts related to industrialization	8	4-6
Recalls facts about industry and technology in Canada	7	3-5
Understands key ideas in economics, history, and geography, specifically those related to economic systems and technological change	11	5-8
Understands underlying values and key issues related to industrial growth	8	4-6
Can identify major issues and recognize differing viewpoints on industrial growth	9	5-7
Can select appropriate questions and apply effective ways of gathering and organizing information	11	6-8
Can analyze, evaluate, and synthesize information related to a topic under study	9	5-7

Target means for each question in the written-response section would be less meaningful than for the multiple-choice section since the total possible score for each question was 5 or less. Consequently, a target mean is set only for the written-response section as a whole, and a target proportion of students achieving "satisfactory" or better is set for Question 5. These regions are given in Figure 7 below.

Figure 7

Target Regions for Grade 9 Social Studies Achievement Test
Written-Response Section

	<u>Number of Marks</u>	<u>Target Region Average Raw Score</u>
Total Written-Response	30	14-19
	<u>Satisfactory Score</u>	<u>% of Students</u>
Question 5 (Resolve and Communicate)	3-5	60-70

To interpret the results for a jurisdiction, the reader should first examine the test specifications, then read the sample questions and the scoring categories and decide whether the target means are appropriate for that jurisdiction. For jurisdictions with fewer than 25 students writing, the targets should be accepted with caution.

Next, the jurisdiction averages on the multiple-choice component should be compared with the target regions and areas of strength and weakness noted. Because the provincial values are of interest, the reader may wish to compare performance with the provincial values. For the total score on the multiple-choice test, jurisdictional averages within four raw score points of the provincial average would indicate no reliable difference between the jurisdiction and the provincial average. For the subtests, two points would indicate a reliable difference.

When comparing the written-response results with the target regions, the reader should keep in mind that the values being compared are proportions of students, not average scores. These provide only one indication of jurisdictional performance, and the reader should also look at the proportion of Category 5 performances that were received from students on Question 5, as well as the proportion of Category 1 performances. (See Table 4, page 28.) No guidelines have been set for Categories 1 and 5, but the provincial results provide a standard for comparison. For these results, a 5% difference is an adequate rule of thumb to use in assessing reliable differences from the provincial results.

Absentee Rates

If more than 10% of the eligible students in a jurisdiction did not write the test, the reported average for that jurisdiction may not accurately represent the jurisdiction averages. Teacher-assigned marks for students who did not write could be compared with teacher-assigned marks for students who did write. If the averages are the same for the two groups, the reported jurisdiction averages are probably representative. If the averages are different, some estimates can be made of what the jurisdiction averages might have been if all students had written the test. Jurisdictions with high absentee rates may wish to contact the Student Evaluation Branch for assistance in estimating their jurisdiction averages.

In examining the test results, the reader must keep in mind that a test score does not indicate why a particular performance occurred, but only that it did occur. After studying the results, the identification of reasons for that performance should be undertaken. There are a variety of factors that should be examined.

1. Student motivation. Consideration should be given to the degree to which students were motivated to perform to their levels of ability.
2. Student ability. While the notion of a target region is designed to take into consideration year-to-year fluctuations in the average ability levels of students, it is possible that a group of students with a particularly high or low average ability may come through a system. This is much more likely to be a factor in small systems than in large ones.

3. Readability. The Achievement Test was designed to a Grade 9 reading level. Jurisdictions should consider the average reading level of their Grade 9 students, as reading levels below Grade 9 will have an effect on test results that will be independent of achievement in social studies.
4. Teaching and curriculum. Consideration should be given to the type of instruction students have received in the jurisdiction and the adequacy of curricular implementation.

There will be other factors that are of importance in particular jurisdictions. School boards wishing to examine further the results in the light of local factors are encouraged to establish their own local interpretation panels.

At a provincial level, some attempt was made to gauge student motivation through a survey administered to the 84 Grade 9 teachers who participated in the scoring of the written-response section. Survey results indicated:

1. Absenteeism for the test was not a problem at the provincial level (99% reported attendance at the test was the same or higher than normal).
2. Motivation approached but was not equal to motivation for a final examination. In general, student attempts at revision, student comments, and other indicators supported this view. These indicators of student motivation show a level of interest adequate for valid test results at the provincial level.

Standard Setting

Another useful statistic for comparison is the percentage of students achieving a "pass" mark on the test. Since tests can vary greatly in their degree of difficulty, a pass is not necessarily 50%. Instead, various standard-setting procedures involving the judgment of experts are used to determine the number of questions a student at a particular level of competence would answer correctly, assuming that he had been exposed to adequate teaching and resources to achieve the curricular objectives. In the case of the Grade 9 Social Studies Achievement Test, a "borderline competent student" was defined as one at the 15th percentile. The Technical Review Committee (with representatives from The Alberta Teachers' Association, the Conference of Alberta School Superintendents, universities, and Alberta Education) used data provided by 84 classroom teachers to establish a pass mark of 40% on the multiple-choice section of the test, or a raw score of 24. This implies that 85% of the students taking the test should achieve a score of 40% or higher on the multiple-choice section, given adequate teaching and resources.

Results

Tables 1 to 5 summarize the results for the test.

The average mark for all students for the complete test was 59.2%. Students achieved an average of 36.6 marks out of 60 on the multiple-choice section, and 16.5 marks out of 30 on the written-response section.

On the 20 multiple-choice questions dealing with Topic A, the provincial average was 12.6; on the 20 Topic B questions, 12.2; and on the 20 Topic C questions, 11.8.

On the written-response questions, students averaged 3.1 marks out of 5 on Question 1, 2.0 marks out of 4 on Question 2, 2.3 marks out of 4 on Question 3, 0.9 marks out of 2 on Question 4, and 8.2 marks out of 15 on Question 5.

Question 5 was scored by judging which of six category descriptions best described each answer. Nearly half (46.9%) of the student responses were placed in the Satisfactory category. The Proficient category received 15.7%, the Exceptional category received 2.7%, the Limited category received 25.8%, and the Poor category received 5.8%. The remaining 3.1% were rated insufficient for categorization.

All figures have been adjusted for the effect of sampling in two jurisdictions by projecting each sampled jurisdiction's results to the total enrolment in that jurisdiction.

Table 1

Total Test Results

	Mean	Standard Deviation
Total	59.2	15.9
Multiple-Choice Section	36.6	10.3
Written-Response Section	16.5	5.4

Table 2

Multiple-Choice Section Results

Reporting Category	Mean	Standard Deviation
Total Multiple Choice	36.6	10.3
1. Topic A: Selected Market Economies	12.6	3.5
2. Topic B: Selected Centrally Planned Economies	12.2	3.8
3. Topic C: Industrialization in Canada	11.8	4.2
4. Basic Facts about Industrialization	5.1	1.6
5. Basic Facts about Canadian Industrialization	4.1	1.7
6. Key Concepts about Economic Systems	6.0	2.3
7. Understanding Underlying Values	5.1	1.9
8. Focus on the Issue	6.2	2.0
9. Research Questions and Data-Gathering	6.9	2.3
10. Analysis, Synthesis, and Evaluation of Data	5.3	2.1

Table 3

Written-Response Section Results

Total	Mean	Standard Deviation
	16.5	5.4
Subtests		
1. Identify Value Conflict	3.1	1.2
2. Recall Knowledge	2.0	1.5
3. Analyze and Evaluate Data	2.3	1.4
4. Synthesize Data	0.9	0.8
5. Extended Response	8.2	3.0

Table 4

Written-Response -- Question 5

Distribution of Student Scores by Category	
Score Category	Frequency in %
5 (EXCEPTIONAL)	2.7
4 (PROFICIENT)	15.7
3 (SATISFACTORY)	46.9
2 (LIMITED)	25.8
1 (POOR)	5.8
0 (INSUFFICIENT)	3.1

Table 5
Item Response Frequencies

Item #	Key	Percentage of All Students Choosing Alternative			
		A	B	C	D
1	C	9	10	77	4
2	B	37	34	17	12
3	C	8	8	41	43
4	D	7	5	6	83
5	D	17	6	10	66
6	C	3	11	66	20
7	A	61	8	18	13
8	D	7	12	5	76
9	C	23	4	58	15
10	B	5	87	4	4
11	B	8	74	11	7
12	B	9	46	27	18
13	D	11	3	5	81
14	C	12	18	46	23
15	C	5	10	70	16
16	A	54	11	24	10
17	C	22	17	60	1
18	D	16	24	9	51
19	B	19	62	13	7
20	D	3	4	21	71
21	D	11	11	4	74
22	A	79	5	6	11
23	C	2	5	85	8
24	C	40	13	43	4
25	A	60	23	9	9
26	D	18	11	14	57
27	A	62	27	6	5
28	B	3	78	14	5
29	A	52	7	15	27
30	C	19	13	43	25
31	B	14	50	13	23
32	B	10	62	12	16
33	D	4	11	18	66
34	D	18	7	35	41
35	C	5	8	76	10
36	B	9	69	16	6
37	B	27	43	12	17
38	A	75	6	14	5
39	B	2	53	9	36
40	C	16	18	48	18
41	D	10	5	34	51
42	D	7	9	13	72
43	D	8	18	9	65
44	A	66	9	13	12
45	D	21	13	17	49

Table 5 (Continued)

Item #	Key	Percentage of All Students Choosing Alternative			
		A	B	C	D
46	B	21	50	20	9
47	A	75	10	10	5
48	A	73	11	10	6
49	A	34	25	25	16
50	B	7	66	11	17
51	A	51	25	12	12
52	C	9	18	65	8
53	A	49	13	18	19
54	B	9	54	31	6
55	D	9	6	16	69
56	C	8	9	74	8
57	C	11	11	58	20
58	D	9	12	17	62
59	B	20	59	11	11
60	A	41	15	15	29

Although performance on different parts of the test shows some variation, these scores are not directly comparable. The sets of items that make up the different parts were not selected to be equal in average level of difficulty. Thus, differences may be due to variations in item difficulty rather than in student performance. In combination with jurisdictional results, however, the norms can be used to detect patterns of relative strength or weakness in achievement.

It must be emphasized that the purpose of the Achievement Testing Program is to produce results valid at the jurisdictional and provincial levels, not at the individual class or student level. Because of restrictions of time and space, the test questions are only a small sample of possible questions for any curricular objective. Thus, individual student scores may vary greatly, depending on the specific questions selected. Accordingly, the achievement tests were designed to evaluate the performance of programs within a jurisdiction or within the province. In addition, conclusions about the performance of any given class should not be made because of the small number of students in most classes and because there is a large variation in the ability of students in different classes.

If the numbers are sufficient, it is meaningful to compare the achievement of groups of students to the provincial averages. For ease of comparison, jurisdiction results are presented in tables that parallel the tables of provincial results. Jurisdictions with fewer than 25 students should exercise caution when comparing results with provincial averages.

Chapter 5

CONCLUDING OBSERVATIONS

In this report, guidelines are provided to assist local boards and administrators in interpreting the Grade 9 Social Studies Achievement Test results for their jurisdictions. Target regions are designed to help local decision-makers assess whether the achievement of their students, when compared with the provincial results, is sufficiently different to be educationally important.

- The provincial mean is 59.2%.
- 87.8% of students achieved a mark of 40% or higher on the test. (40% is considered the minimum acceptable standard and 85% of students should be at or above this standard, given adequate teaching and resources.)
- 71.4% achieved a mark of 50% or higher and 39.3% achieved 65% or higher.
- 10.8% achieved a mark of 80% or higher.
- The provincial mean on the written-response section of the test was 16.5 (55.0%).
- 67.5% of the students produced acceptable or better compositions for question 5, another 25.8% showed limited achievement, and 5.8% produced poor responses.
- At the provincial level, student achievement was above standard on the multiple-choice section of the test and below standard on the written-response section. Overall, achievement on the Grade 9 Social Studies Achievement Test was at an acceptable standard.